# In Search of the Name

W. Dean Bidgood, Jr., M.D., Departments of Biomedical Engineering and Radiology, University of North Carolina at Chapel Hill, North Carolina
Wayne R. Tracy, M.S., Cerner Corporation, Kansas City, Missouri

## ABSTRACT

*Existing communications standards represent person name, date, time, and other ubiquitous attributes in various incompatible formats. The electronic medical record requires convergence of diverse representational systems toward a single communications standard or a harmonized group of standards. The obstacles to convergence include disparities in semantic definition, syntax, and communications protocols. To facilitate harmonization of existing standards, the message standards developers subcommittee of the ANSI HISPP (American National Standards Institute Healthcare Informatics Standards Planning Panel) has defined a set of common data types to facilitate semantic convergence. The authors present the general method used to develop the common data types. The derivation of the person name common data type is presented in detail. A general semantic model of the person name attribute is developed from observations of international usage conventions. A superset of the person name formats of the ACR-NEMA, ASTM, HL7, NCPDP, MEDIX, and X-12 standards is taken as the provisional starting point for a common data type definition. The convergence superset is compared with the general semantic model. Highly specialized and/or infrequently encountered sub components of the general model are combined into component complexes, thereby defining mappings to less rigorous representations. The ANSI HISPP common data types are specified for use in a demonstration of a prototype object-oriented HL7-DICOM HIS/PACS interface (between hospital information systems and imaging systems) at the 1993 Annual Meeting of the Radiological Society of North America.*

## INTRODUCTION

Accurate communication of a person's name is essential in healthcare, even though the person name itself is not a unique identifier. While a global unique identification scheme theoretically would be desirable, none is currently implemented. Therefore, a person's name remains a vital component of algorithmic searches that utilize a set of demographic items to distinguish among individuals. Increased precision in recording names, along with careful attention to "aliases", maiden names, and other names facilitates the production of a longitudinal record. Linkage across multiple encounters offers the opportunity to understand a person's health status in a continuum rather than in an isolated episode.

Domestic United States healthcare providers often are confronted with significant populations of foreign born individuals whose names have been either imprecisely or incorrectly represented, limiting access to antecedent records. Efficiencies both in the interview dialog and in the recording of historical information surrounding the individual's change in health status accrue if the precision of the electronic record is improved. In the United States, the typical computer data definition of person name lacks essential robustness. This is particularly true, for example, for names of Hispanic origin and for hybrid and compound names that are being created with increasing frequency. Common semantic understanding is needed to support electronic sharing of patient records across systems using different standards as well as across national boundaries. A general semantic model of the composition of person name is necessary, along with a mapping strategy to allow less rigorous instantiations of the person name attribute in computer databases. We present a method of definition of a common data type as a template for further work to achieve semantic convergence of communications standards in medical informatics and to develop expert systems for automatic parsing of sub components of person names.

## METHOD

The definition of a common data type requires first an analysis of the semantic composition of the real-world attribute represented by the data type. The analysis, is based on principles of

0195-4210/92/$5.00 

abstract data type development. In view of the increasing need for international cooperation in health care, one must consider various international perspectives as well as typical Anglo-Saxon usage. High and low priority semantic components and sub components are identified and recorded. The next step is compilation and reconciliation of the representations of the attribute by the pertinent communications standards. A superset of the existing representational systems is taken as the provisional starting point for a common data type definition. The convergence superset is compared with the general semantic model. A practical compromise is achieved by combining highly specialized and/or infrequently encountered sub components into component complexes to facilitate implementation. Thus, a general semantic model is defined, together with mappings to less rigorous representations. The utility of the proposed common data type is determined from observations of functional integrity in prototype applications and from the degree of acceptance by system implementors.

## GENERAL SEMANTIC MODEL OF PERSON NAME

Table 1. General semantic model of person name

1. Surname complex
    - Family name group
        - Paternal family name set
            - { Family name prefix
            - Family name }**
        - Maternal family name set
            - { Family name prefix
            - Family name }**
    - Composite surname group
    - Synthetic surname group
2. Given name complex
    - Family given name group
    - Place of origin
    - Deity or patron saint group
    - Religion
    - Caste or class
    - Physical or personal attribute group
    - Other groups*
3. Prefix complex
    - Deferential title prefix group
    - Gender, age, marital status prefix
    - Professional prefix group***
4. Suffix complex*
    - Hereditary order suffix
    - Professional suffix group***
    - Degree group

The above classification of person name components and sub components is based on qualitative analysis of anecdotal data from past experience and interviews with individuals from a variety of ethnic origins. In Table 1, the term "complex" denotes a superclass of diverse but related classes (groups). A "group" denotes a homogeneous class of entities or sets of entities. Those components designated as "complex", "group", and "set" may contain multiple values. The other components are believed to have only a single value at any one time for a given person.

**Examples:**

1.Family name prefix: *dalla, di, die, el, von, viuda de, and vda. de.* Multiple pairs of family name prefix and family name are known to occur in a sequence of surnames. Therefore, the family name sequence must consist of alternating family name prefix and family name sub components. Any given surname prefix sub component (or all of them) may have a value or may be a null set.

2. Given name complex: *Franklin Delano, Juan Pedro de La Madrid, Moon Unit.* Given names are highly variable. They may consist of family surnames or given names, local deities, place of birth, or any random name chosen by the parents or others. The first entity of the given name complex maps to "first name" in typical United States usage. The secondary entity or entities map to "middle name(s)" in typical United States usage.

3. Prefix complex: Deferential title sub component: *Right Honorable, Sir, Don, Doña.* Gender, age, marital status prefix sub component: *Miss, Ms, Mr, Señora, Señorita, Mlle.* Professional prefix sub component: *Dr,*

---

* *The authors recognize that this is a very partial representation of name structures. Our intention is to reassess the general model as we encounter additional examples from diverse ethnic and cultural usage.*

** *Brackets "{}" indicate that the two components of the set can appear in a regular repeating alternating series.*

*** *Certain attributes, such as military rank, may appear with identical values in either a prefix complex or a suffix complex, but not in both simultaneously.*

*Rev, Pres, Captain, Sergeant.* Note that at least three semantic dimensions (gender, age, and marital status) are implicit in the common prefixes of address (Ms, Mr, Master, Miss, Mrs).

4. Suffix complex: Hereditary order sub component: *Jr, Sr, III, or IV.* Professional suffix sub component (order, society or honorary membership, military rank, branch of service, modifier): *FACS, FACR, Captain, Sergeant, USAMC, USAF, RET.* Degree sub component: *BS, MSEE, MPH, JD, MD, M.Div.*

## COMPILATION OF EXISTING STANDARDS

Table 2. ASTM PN Person Name [1]

| | |
|---|---|
| 1. | Last Name |
| 2. | First Name |
| 3. | Middle name or initial |
| 4. | Name suffix, e.g. Jr. or III |
| 5. | Prefix title, e.g. Dr, Mr, Ms |
| 6. | Degree, e.g. pd, DDS, MD |

Table 3: ACR-NEMA Version 2.0, Group 0010 Patient name data elements [2]

| | |
|---|---|
| (0010,0010) | Patient's name |
| (0010,1001) | Other patient names |
| (0010,1005) | Patient's Maiden Name |
| (0010,1060) | Patient's Mother's Maiden Name |
| (0010,1080) | Military Rank |

ASTM (Table 2) specifies six components for person name. Neither Medix nor NCPDP specifies any person name sub components that are not included in the ASTM specification. ACR-NEMA version 2.0 (Table 3) specifies only a single patient name component, but defines several elements related to person name. The Japanese Image Save and Carry (IS&C) Standard, based on ACR-NEMA version 2.0, includes a single data element for Patient Name (0010,0010) [3]. This is encoded by tag-length-value (as a variable length field). The HL7 standard is similar to the ASTM Person Name [4]. X12N 835 is similar to ASTM with exception of the omission of component 6: Degree (e.g. Ph.D., M.S., M.D.).

## COMMON USAGE

The following examples describe specific cultural or national usage. They are the source material for the more abstract representation presented in Table 1:

Case 1: In the United States and other English-speaking countries, person names consist of one or more given names. The first is designated the "first name" and the second is designated the "middle name". Traditionally, the father's "last name" becomes the child's family name ("last name"). Traditionally, a woman assumes the family name of her husband when she marries. Wives in increasing numbers are retaining their family names. Compound maternal and paternal names (in either order) are commonplace. Adopted children have been given the family name of one adoptive parent, or compound names consisting of the family names of both adoptive parents (or partners, in some cases). Other couples have legally defined their children's family name as a composite of the parents' family names (e.g. Skybetter, formed of Skylar and Ledbetter). Others have created completely synthetic new family names for their children). Even more confusingly, others rotate their family names periodically [5]. While variability is considerable, American and British names can be represented by a 4 component structure consisting of family name complex, given name complex, prefix complex, and suffix complex, with unrestricted iterations of any component (see Table 1).

Case 2: In conventional Spanish and Latin American usage, children have one or more given names and a compound family name consisting of a paternal component and a maternal component. Surname prefixes commonly are present. It is common to enumerate multiple generations of maternal and paternal family names in a person's name under some circumstances. One or more given names are present. Composite contractions of paternal and maternal names occurs (e.g. after marriage, Loperena Menendez becoming Lopez).

Case 3: In conventional German usage, a child has a single given name (occasionally more than one). The child typically has taken the father's family name. However, in recent years, almost

any possible last name structure has been used (as in the United States and other English speaking countries). For example, a husband can choose to use a compound name consisting of his wife's family name followed by his own family name (e.g. Mr. Gauger marries Miss Schmidt and he changes his name to Mr. Schmidt-Gauger). It is not uncommon in married couples for the husband to retain his family name while the wife retains hers or assumes a compound family name (with either the husband's or wife's family name as the first component). At age 18, regardless of the structure of the birth name, a child has the opportunity to change his or her family name.

Case 4: Name structure varies widely in India. Typically, one or two given names are used. A family name is appended. Uncommonly, as many as five given names may be present. Given names may be derived from the names of family elders, the god or goddess worshipped, ancestral heritage, a location, or an adjective describing an attribute that the parents desire the child to have (e.g. veera, meaning bravery).

Case 5: Name structure in Korea is relatively simple. Given names typically have one or two syllables. Nearly all family names have only one syllable. A small minority have two syllables. In many cases, brothers carry the same syllable in their names. For many families, an entire generation share the same syllable in their names. Traditionally, the father's family name becomes the children's family name. Two exceptions exist: 1) When an unmarried woman adopts a child and 2) when an unmarried couple have a baby and the man denies fatherhood or refuses to accept the child, the child received the woman's family name.

Case 6: Vietnamese names (from the region around Saigon) typically consist of three components at birth. The first given name component is unrestricted. For a daughter, the second given name component is traditionally her mother's family name. The third component is the father's family name. The usual written order is components three, two, then one. The youngest daughter sometimes has an additional given name (however, some daughters have only two names). Boys' names typically have three components: Two given names (one "free" choice plus the father's middle name) and the father's family name. No divorce or remarriage is permitted in the traditional setting. A woman does not change her name when she marries.

## CONVERGENCE SUPERSET

The ASTM E-1238-91 standard person name data type essentially is a superset of the other standards for name representation in medical informatics. The ASTM "middle name" component is redundant semantically, but is retained because of the overwhelming implementation of "middle name" in Anglo-Saxon name representation. In most cases in the United States the "middle name" is the second instance of the "given name complex" of the general model

The "degree" component of ASTM (a sub component of the "suffix complex" of the general semantic model) is subsumed by the "suffix" component of the common data type For practical implementation, all sub components of the general semantic structure are aggregated and mapped into the corresponding major components. Note: The "surname complex" of the general model is designated the "family name" in the ANSI HISPP person name common data type structure.

Table 4. ANSI HISPP Person Name Common Data type Structure

1. Family name (equivalent to ASTM and HL7 last name)
2. Given name (equivalent to ASTM and HL7 first name)
3. Middle name
4. Prefix
5. Suffix

1). Any of the above may repeat.
2). In typical American and European usage the first occurrence of the "given name" would represent the "first name". The second and subsequent occurrences of the "given name" would typically be treated as "middle name or names". The implementor should remain mindful that translations to and from the typical usage of "middle name" may be required.
3). The "degree" component is subsumed by the "suffix" component.
4). [Syntax note]: Interior null components (e.g. no middle name) must be indicated as such (e.g. "no known value" or "value does not exist").

Trailing null components may be ignored (e.g. one may assume that no value exists).
5). [Syntax note] Where multiple iterations of a sub component exist, they should be listed in order from greatest to least significance (when "significance" is defined and/or known).
6). [Syntax note] Name entries are encoded as literal text strings for simplicity. Note that this compromise from the general semantic model blurs the distinction between surnames and surname prefixes. [6]

## UTILITY OF THE ANSI HISPP COMMON DATA TYPES

The common data types were a major incentive for cooperation between the HL7 and ACR-NEMA organizations on an international demonstration of HIS/PACS systems integration for the 1993 annual meeting of the Radiological Society of North America. ACR-NEMA adopted the initial set of ANSI HISPP common data types to ensure compatibility of DICOM version 3.0 with other major standards. [7,8]

## CONCLUSION

The need for a common semantic understanding of the data elements of the computer based medical record is compelling. The ANSI HISPP common data types provide essential common ground for medical informatics standards developers. The person name common data type is sufficiently comprehensive to accommodate American and British names, a variety of international names, and hybrid naming formats. A compromise five-component structure (allowing repetition within each component) significantly reduces the data encoding complexity for routine applications (versus the complexity of the general semantic model) and makes allowances for backwards compatibility with data and message structures conforming to older versions of the major standards. Future applications, such as rule-based systems for automatic decomposition of person names into standard parts, will further specialize existing systems by distinguishing the grouped sub components of the general four-component semantic model. Comprehensive computerized patient records including a database of current name, birth name, and previous name instances (as an alias recursion) in standard format will be amenable to processing by inference systems utilizing a knowledge base of naming conventions. Works in progress include enhancement of the general semantic model by evaluation of other ethnic and cultural patterns and verification of the integrity of the person name common data type in prototype applications. Vendors of hospital information systems and imaging systems will test the common data type concept in an international demonstration of HIS/PACS system integration at the December, 1993, annual meeting of the Radiological Society of North America. Imaging system developers have responded favorably to the conversion from an oversimplified single-field name structure (used in ACR-NEMA version 2.0 implementations) to the more precise ANSI HISPP common data type format (specified in ACR-NEMA DICOM version 3.0).

## References


1. E 1238-91 Standard Specification For Transferring Clinical Observations Between Independent Computer Systems. Revision 4.2.1. C. J. McDonald, chairman. ASTM Committee E-31 on Computerized Systems, Indianapolis, IN, February 11, 1992.
2. ACR-NEMA V3.0 - Digital Imaging And Communications In Medicine (DICOM) - Part 6: Data Dictionary. Draft For Informal Review, Version 0.6. S. C. Horii, editor. National Electrical Manufacturers Association, Washington, D.C., November 18, 1992.
3. Image Save and Carry (IS&C) Standard Version 1.0. Data Format. (English language version). IS&C Committee, Working Group 2. Tokyo, Japan, February 21, 1992.
4. Health Industry Level 7 Interface Standards, Version 2.1. Health Level Seven, Inc., Ann Arbor, MI, June 30, 1990.
5. Kaplan, Janice L. A new, complicated twist in naming children. The Gainesville Sun. Gainesville, FL. January 31, 1993, p.5D.
6. ANSI Healthcare Informatics Standards Planning Panel Common Data Types. W.D. Bidgood, Jr. and W.R. Tracy. American National Standards Institute. June 18, 1993.
7. W.D. Bidgood, Jr., S.C. Horii. Introduction to the ACR-NEMA DICOM Standard. Radiographics 12:345-355, 1992.
8. The ACR-NEMA Digital Imaging and Communications Standard (Version 3.0; Part 5). National Electrical Manufacturers Association.. Washington, D.C. 1993.